## Mikor volt az egyes digitális, bibliográfiai jellegű adatbázisok aranykora? avagy esettanulmány hasonlóságelemzésekre alapozott, konzisztens (vállalati) elemzési kultúra növelése érdekében

Szemesiné Cseh Kata, KJE

Balatoni Regionális Történeti Kutatóintézet, Könyvtár és Kálmán Imre Emlékház könyvtárosa

## Kivonat

A Google Trends-en futtatott lekérdezések alapján lehetséges a digitális, bibliográfiai jellegű adatbázisok aranykorának levezetése. A Matarka kivételével a többi adatbázis, mint MTMT, ARCANUM, EPA, OSZK, SCHOLAR, academia.edu, Researchgate, archive.org, MEK, alapvetően egyensúlyi állapotokat mutat, a 2004.0.1.01.-2023.06.23. között éves bontásban vizsgálva az adatokat, hasonlóságelemzések sorozatára támaszkodva. Néhány szolgáltatás (mint az MTMT, vagy az OSZK és a MEK) két-két év között mutat ellentmondásokat. Ezek az ellentmondások, alacsony számuk miatt valószínűsíthetően a Covidhoz köthetőek, mert rövid hatásmechanizmusúak. A MATARKA esetében, a 2004-2009-ig tartó semleges időszakot követi az aranykor, amely 2010-től 2016-ig tartott, ezt követi egy negatív időszak, amely napjainkban is tart. Az idő múlása is hibátlanul modellezhető az adatok alapján, amely a fenti eredményeket egy fajta speciális konzisztencia-keretbe foglaltan erősíti meg.

## Kulcsszavak

Digitális könyvtár, MATARKA, ARCANUM, MTMT(Magyar Tudományos Művek Tára), EPA (Elektronikus Periodika Adatbázis), OSZK, SCHOLAR, academia.edu, researchgate, archive.org, MEK(Magyar Elektronikus Könyvtár)

## Idegen nyelven is átadandó rétegek

Title: When was the golden age of some bibliographic databases? or a case study based on similarity analysis, to increase a consistent (corporate) culture.

Abstract: Based on the queries run on Google Trends, it is possible to infer the golden age of digital, bibliographic databases. With the exception of Matarka, the other databases, such as MTMT, ARCANUM, EPA, OSZK, SCHOLAR, academia.edu, researchgate, archive.org, MEK, show basically equilibrium states, looking at the data in a yearly breakdown from 01.0.1.2004 to 23.06.2023 (date of the query 23.06.2023), based on a series of similarity analyses.

Some services (such as MTMT, or OSZK and MEK) show inconsistencies between two or three years. These inconsistencies, due to their low number, are likely to be related to Covid because they have a short impact mechanism. In the case of MATARKA, a neutral period from 2004-2009 was followed by a golden age from 2010 to 2016, followed by a negative period that continues today.
The passage of time can also be modelled flawlessly from the data, which confirms the above results in a kind of special consistency framework.

Keywords: digital library, MATARKA, ARCANUM, MTMT (Magyar Tudományos Művek Tára), EPA (Elektronikus Periodika Adatbázis), OSZK, SCHOLAR, academia.edu, researchgate, archive.org, MEK (Magyar Elektronikus Könyvtár)

## Bevezetés

Napjainkban egyre nagyobb igény van a dokumentumok digitális változatára. A kiadók nemcsak a nyomtatott dokumentumokat, de ezek digitális változatait is elérhetővé teszik. Ennek oka a nyomtatási költségek megnövekedése, ennek következtében a nyomtatott dokumentumok ára is emelkedik, valamint növekszik a digitális dokumentumok iránti igény. A könyvtárak a nyomtatott dokumentumok csak egy részét képesek megvásárolni az egyre szűkülő anyagi lehetőségeik miatt. A kultúrára szánt források csökkennek. A digitális dokumentumok vásárlására is egyre kevesebb könyvtárnak van lehetősége. A költségeket tekintve ugyan kedvezőbb árakkal számolhatnának, de a tárolási és kölcsönzési feltételek bonyolultsága miatt egyelőre maradnak a hagyományos nyomtatott forma mellett

### Célok

A célom, hogy választ találjak a digitális, főként csak bibliográfiai adatokat tartalmazó adatbázisok létjogosultságára, a feléjük irányuló érdeklődés racionalitására, ill. ingadozására.

### Feladatok

Megvizsgálni a különböző digitális adatbázisokra futtatott keresések alapján, hogy van-e értelme még azokat a platformokat életben tartani, amelyek csak a dokumentumok bibliográfiai adatait tartalmazzák.

### Motivációk

Egy-egy feladat elvégzésében problémát okozhat az elvégzéshez szükséges dokumentum digitális formájának hiánya. Hiába vannak meg a könyv, dokumentum bibliográfiai adatait egy adott adatbázisban, a könyvhöz, dokumentumhoz való hozzáférés akadályokba ütközhet. Amennyiben valaki könyvtári tagsággal rendelkezik könyvtárközi kölcsönzésben megkérheti, aminek a költsége – postaköltség és csomagolási költség - gyakran túlszárnyalja a könyv árát is. Itt nemcsak régi kiadású dokumentumokról van szó, de az újabb megjelenéseknél – főként szakkönyvek esetében – is nehézkessé válhat. Ami meglepő, hogy igen régi dokumentumokat lehet fellelni különböző meghíváson alapuló adatbázisokban. Na, igen ezeknél az adatbázisoknál a szerzői jog nem számít releváns tényezőnek., ahogy mondani szokták, „A szükség nagy úr” alapján a fekete és fehérnél rögtön jobban tetszik a szürke. Azonban ettől függetlenül nagyon sokan szeretnének hozzáférni ezekhez a dokumentumokhoz, még ha a hozzáférés csak ideiglenes is.

### Célcsoportok

Nincs meghatározott célcsoport. A dokumentumok hozzáférhetőségének kérdése mindenki számára fontos. Ha célcsoportot kellene meghatároznom főként a 10-50 éves korosztályt választanám. A fiataloknak a tanuláshoz van a digitális anyagokra szükségük, az idősebb korosztály pedig, akik már számítógépet és internetet, sőt okostelefont használnak, főként a hírek, könyvek iránt fordítanak érdeklődést.

### Hasznosság

A napjainkra jellemző időhiány, hisz az „idő pénz”, ráadásul a világ felgyorsult, mindenki minél több tartalommal szeretne a legrövidebb időn belül találkozni. Az nem kérdés, hogy megtalálhatjuk a keresett dokumentum adatait, de vajon mennyi időbe telik, mire a tényleges dokumentumot a kezünkbe vesszük. Egy óra, vagy egy hét? Ezért kezdtem vizsgálni az adatbázisokat, hogy mennyire van értelme ma a bibliográfiai adatoknak, digitális dokumentum nélkül.

## Szakirodalmi/saját előzmények

A digitális bibliográfiai adatvagyonok kezelésének aranykorával a szakirodalom és az érintettek érdemen nem foglalkoznak/foglalkoztak:

* <https://abtk.hu/ismerettar/ismeretterjesztes/1446-bibliografiai-adatbazis-az-oktatas-es-a-kutatas-szolgalataban>
* <http://epa.oszk.hu/00600/00691/00078/pdf/mtud_2010_06_684-693.pdf>
* <https://matarka.hu/index.php>
* <https://www.arcanum.com/hu/>
* <https://m2.mtmt.hu/gui2/?type=topic&mode=browse&sel=topic10119>
* <https://epa.hu/>
* <https://mek.oszk.hu/>

### A probléma/jelenség története

Bizonyos dokumentumok digitális fellelhetőségének hiánya. Pár évvel ezelőtt találkoztam egy igen jelentős problémával, miszerint egy bizonyos dokumentum adatait ugyan megtaláltam egy, vagy több adatbázisban, de magához a dokumentumhoz, vagy annak másolatához nem jutottam hozzá, csak meghatározott idő múlva, rosszabb esetben költségek mellett. Könyvtárosként rendelkezem azzal az előnnyel, hogy költségek nélkül megkapom a keresett dokumentumot könyvtárközi kölcsönzésbe, de az az időfaktor akkor is fontos tényező. Ezért kezdtem meg a vizsgálatot a Google Trends segítségével, hogy vajon még most is sokan felkeresik-e a fenti digitális adatbázisokat és az adatbázisok tartalmában történt-e változás.

### A probléma/jelenség aktuális állapota

A dokumentumok teljességgel való – lehetőleg azonnali - fellelhetősége kulcskérdés a mai felgyorsult, és stresszes világban. Az adatokra azonnal lenne szükségünk, ami nem minden esetben valósul meg. Az alábbi táblázatban a digitális adatbázisokra, az elmúlt 20 éveben lefuttatott keresések adatai láthatóak (vö. 1. ábra).

### A probléma jelenség adatvagyona



1. Ábra: Az éves bontású adatvagyon (forrás: <https://miau.my-x.hu/miau/302/mtmt_matarka_epa_oszk_.xlsx>)



1. Ábra: Az idő, mint olyan modellezhetősége (forrás: <https://miau.my-x.hu/miau/302/mtmt_matarka_epa_oszk_.xlsx>)

A 2. ábra üzenete egyszerű: az idő múlása alapvetően értelmezhető (vö. <https://miau.my-x.hu/miau/303/futurology_physics_time.docx>), vagyis az adatvagyon az összes objektumra nézve alapvetően stabil.



1. Ábra: Az ARCANUM tény vs. becslés viszonya (forrás: <https://miau.my-x.hu/miau/303/futurology_physics_time.docx>)

Az Arcanum-nál lefuttatott becslésnél látszik, hogy nincs eltérés a valódi és a becsült adatok között. Az ARCANUM kapcsán a Google Trends érdeklődés tehát úgy alakult, ahogy azt racionálisan el lehetett várni a többi objektum tükrében.



1. Ábra: A MATARKA aranykora (forrás: <https://miau.my-x.hu/miau/303/futurology_physics_time.docx>)

Itt (vö. 4. ábra) jelenik meg az aranykor nevű jelenség. Ez azt jelenti, hogy a valós adatok 2009-2015 között meghaladták a becsült adatokat. Tehát sokkal többen kerestek rá a MATARKA kulcsszóra, mint ahogy a program az adatokból becsülte, mint ahogy az elvárható lett volna. 2016-tól napjainkig a MATARKA iránti érdeklődés rel. gyenge, s egyre nő a becslés szerint elvárható és a valóságban realizált értékek távolsága, ami vészjelzés a MATARKA (mint projekt, vállalkozás) részére – feltételezve, hogy hasonló „iparági” elemzéseket egy vállalati információs rendszeren belül mindenki végez, aki a jövőbeni fenntarthatóságot fontosnak tartja.

## Potenciális megoldási alternatívák

Az elmúlt időszakban számottevően sok tartalmat töltöttek fel teljesség igényével az adatbázisokban – ezért is kellett az idő múlását, mint X-tényezőt minden esetben figyelembe venni, ill. az időt magát is ellenőrző jelleggel modellezni. De ez még mindig kevés. Még mindig maradtak olyan dokumentumok, amelyeknek csak a bibliográfiai adatai találhatóak meg. Példaként említem a MATARKA adatbázist, amelyben több mint 50 könyvtár adatbázisát használja. Egy probléma van csak vele. A MATARKA több mint 1800 folyóiratot tart számon, de ezen folyóiratok cikkeinek mindössze **30%-hoz** lehet a teljesség igényével hozzáférni. Ez az évente megjelenő cikkek számát tekintve elenyésző. És a külföldi folyóiratok fordításai egy külön probléma. Csak induljunk ki a külföldi, orvosi szaklapokból. Sajnálatos módon az orvosok elenyésző százaléka beszél idegen nyelvet, - lévén elöregedett az orvostársadalom – így csak kevesen tudnak igazán lépést tartani az újításokkal.

## Adatok és módszerek

### Saját adatvagyon

Lekereséseket végeztem a Google Trends-en a MATARKA, Arcanum, MTMT, EPA, OSZK, SCHOLAR, ACADEMIA.EDU, RESEARCHGATE, MEK és ARCHIVE.ORG kulcsszavakra. Időintervallumként 2004-2023 között, területi intervallum Magyarország, valamint nincs kategória megjelölés és csak az internetes keresések érdekeltek. Az eredményeket összegeztem. És átlagot számoltam évekre a keresésekből (vö. 5. ábra):



1. Ábra: A nyersadatok (forrás: <https://miau.my-x.hu/miau/303/futurology_physics_time.docx>)

### Saját módszertan

A kapott nyers adatokat előbb éves aggregációvá alakítottam, majd ezen adatokat sorszámoztam oszloponként, hogy az adott évben, a többi évhez képest hányadik az évek sorában. Attól függően, hogy melyik adatbázisra voltam kíváncsi a többit bemásoltam a COCO rendszerbe a hasonlóságelemzéshez (vö. 6. ábra - <https://miau.my-x.hu/myx-free/>). Az alábbi táblázatban megjelöltem a becslést és a számolt adatokat. A két oszlopból készítettem vonaldiagramot a lekereséseknél (vö. 7. ábra):



1. Ábra: Egy adott objektum modellje: input OAM és lépcsős függvény (forrás: <https://miau.my-x.hu/miau/303/futurology_physics_time.docx>)



1. ábra: Az egyes objektum tény vs. becslés állapotai (forrás: <https://miau.my-x.hu/miau/303/futurology_physics_time.docx>

## Eredmények

A grafikonokból megállapítható, hogy a legtöbb objektum iránti érdeklődés úgy viselkedik, ahogy elvárható. A többi eredmény, mint a MATARKA vagy az MTMT grafikonján látszik, hogy minkettőnek volt olyan időszaka, amikor a becslések alulmaradtak a valódi keresésekhez képest. Az MTMT grafikonján látható, hogy nő a keresések száma, míg a MATARKA adatbázisánál ez csökkenő tendenciát mutat.

## Jelenlegi állapotok

Megnéztem mindkét adatbázist tüzetesebben. Valóban többé-kevésbé az összes folyóirat cikkeinek bibliográfiai adatait tartalmazza, legyen szó magyar vagy külföldi folyóiratról. Ugyanakkor a probléma számomra változatlanul fennáll, vagyis hiába van odaírva, hogy teljes dokumentum. Rákattintva nem nyitja meg a dokumentumot, vagy hibát ír, csak a bibliográfiai adatok vannak meg, melyik folyóiratban található meg.

## Következtetések

 Következtetésként az elmúlt évekre visszatekintve még mindig nincs elég teljes értékű digitális tartalom az adatbázisokban. Tudom, hogy humán erőforrásra lenne szükség, akik tartalmakat feldolgozzák és ingyenesen vagy térítés ellenében hozzáférhetővé teszik. Azonban ahogy már említettem itt az időtényező az egyik legfontosabb része a kereséseknek. Saját tapasztalatból tudok csak meríteni: amikor a harmadik URL-re irányít át az oldal és még mindig nem tudok a teljes cikkhez, vagy tanulmányhoz hozzáférni, az frusztráló tud lenni.

## Jövőkép

A jövőt tekintve nem vagyok bizakodó. Még sok évnek kell eltelnie ahhoz, hogy a publikációk azonnal hozzáférhetőek legyenek. Abba, hogy a publikációkat visszamenőleg is digitalizálják már nem is reménykedem. Egyedül az Arcanum adatbázisa az, amelyik folyamatosan frissül, bővül. Több mint 4500 folyóiratra és könyvre, valamint történelmi térképre lehet keresni, bár ez térítéses. Úgy gondolom, hogy a két adatbázist, vagyis a MATARKA-t és az Arcanum-ot egyesíteni kellene. Hiszen van olyan folyóirat amelyik csak az egyikben fellelhető. A lefuttatott keresések alapján kiderült számomra. hogy nem keresik ezeket az adatbázisokat annyira, mint lehetne. Ha az üzemeltetők jobban törekednének a tartalmak teljes digitalizálására, valószínűnek tartom, hogy megnőne az adatbázisok iránti keresések száma. Hiszen akkor van értelme fenntartani egy ilyen adatbázist, ha keresik, és keresnek benne. Azonban a jelenlegi kép azt mutatja, hogy az igények kielégítéséhez még van hova fejlődni. Mindenki, aki az adatbázis használja, már avval a tudattal futtatja majd a kereséseket, hogy tisztában van vele, hogy a keresés a teljes dokumentum találatával zárul.

## Tartalomjegyzék

[Mikor volt az egyes digitális, bibliográfiai jellegű adatbázisok aranykora? avagy esettanulmány hasonlóságelemzésekre alapozott, konzisztens (vállalati) elemzési kultúra növelése érdekében 1](#_Toc139831514)

[Kivonat 1](#_Toc139831515)

[Kulcsszavak 1](#_Toc139831516)

[Idegen nyelven is átadandó rétegek 1](#_Toc139831517)

[Bevezetés 1](#_Toc139831518)

[Célok 2](#_Toc139831519)

[Feladatok 2](#_Toc139831520)

[Motivációk 2](#_Toc139831521)

[Célcsoportok 2](#_Toc139831522)

[Hasznosság 2](#_Toc139831523)

[Szakirodalmi/saját előzmények 2](#_Toc139831524)

[A probléma/jelenség története 3](#_Toc139831525)

[A probléma/jelenség aktuális állapota 3](#_Toc139831526)

[A probléma jelenség adatvagyona 3](#_Toc139831527)

[Potenciális megoldási alternatívák 5](#_Toc139831528)

[Adatok és módszerek 5](#_Toc139831529)

[Saját adatvagyon 5](#_Toc139831530)

[Saját módszertan 6](#_Toc139831531)

[Eredmények 8](#_Toc139831532)

[Jelenlegi állapotok 8](#_Toc139831533)

[Következtetések 8](#_Toc139831534)

[Jövőkép 8](#_Toc139831535)

[Tartalomjegyzék 9](#_Toc139831536)

[Hivatkozások 9](#_Toc139831537)

## Hivatkozások

…szövegközben…